DOCUMENT RESUME

ED 252 561 TM 850 030

AUTHOR Baker, Eva; Herman, Joan
TITLE Policy Studies: Teacher Competency Assessment and R&D Needs in Educational Testing and Evaluation. Research into Practice Project.
INSTITUTION California Univ., Los Angeles. Center for the Study of Evaluation.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
PUB DATE Nov 84
GRANT NIE-G-84-0112-P4
NOTE 25p.
PUB TYPE Reports - Descriptive (141)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Cost Effectiveness; Educational Policy; Educational Research; *Educational Testing; Elementary Secondary Education; Evaluation Criteria; *Evaluation Methods; Evaluation Needs; Information Systems; *Minimum Competency Testing; *Research and Development; *Research Needs; Teacher Education; *Teacher Evaluation

ABSTRACT

This report summarizes the results of two educational policy studies: (1) an analysis of current problems in teacher competency testing; and (2) an exploration of educational testing and evaluation research and development needs. In the first study, information on teacher competency testing was gathered from a literature review and meetings with representatives from seven educational organizations. The problems with several methods of evaluating teacher coopetency are discussed: the observational approach; outcomes approaches (student achievement); and teacher readiness. It is suggested that both technical and professional approaches using multiple criteria for judging teacher competency be pursued. The second study summarizes the views of prominent scholars at a meeting on educational testing and evaluation research priorities. Research is needed on the concept of validity; new kinds of testing (higher order skills, school curriculum, and diagnostic testing); comprehensive information systems; the effects of current testing; and cost-efficient analyses for educational decision-making. (BS)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED252561

DELIVERABLE - NOVEMBER 1984

RESEARCH INTO PRACTICE PROJECT

Policy Studies: Teacher Competency Assessment and R&D Needs in Educational Testing and Evaluation

Eva Baker

Joan Herman

Project Directors

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Grant Number

NIE-G-84-0112, P4

TM 850 030

Center for the Study of Evaluation

Graduate School of Education

University of California, Los Angeles

The project presented or reported herein was performed pursuant to a grant from the National Institute of Education, Department of Education. However, the opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by the National Institute of Education should be inferred.

Introduction

This report reflects CSE's continuing interest in identifying and analyzing issues of importance to educational testing and evaluation and to the formulation of educational policy. It summarizes the results of two policy study activities: An analysis of current problems in teacher competency testing and an exploration of R&D needs in the field of educational testing and evaluation.

The analysis of issues and problems in devising teacher competency system is the result of a two stage process. During the first stage, a literature review was conducted. Meetings were then held with representatives from the National Education Association - Instruction and Professional Development, the American Federation of Teachers, the Interinstitutional Conference of the American Educational Research Association, the Pennsylvania School Study Council, the U.S. Department of Defense Dependent Schools - Evaluation Division, the Council of Chief State School Officers, and the Office of Technology Assessment to get their perspectives on the problem. The enclosed analysis is preparatory to conduct a conference on teacher competency testing in 1985, intended to provide a forum for sharing ideas among those in the research and policy making communities and to explore potential solutions to existing problems.

The identification of R&D needs in educational testing and evaluation reflects CSE's longstanding commitment to setting the research agenda for the field and highlighting areas needing additional support. The paper summarizes the results of a working

meeting held in Boulder, Colorado devoted to these topics. The meeting, held on August 30, 1984, assembled prominant scholars in the field to give their views on research priorities.

Teacher Competency Assessment: Issues and Agendas

Eva L. Baker

Problem

Of central interest in the evaluation of the educational system is the effectiveness of teachers. This interest has been demonstrated by regulations and legislation designed to attempt to grasp more firmly the quality of teachers through planned assessments of their productivity. Certainly, initiatives for merit pay, whereby master or specially competent teachers would be rewarded, represent a logical extension of student testing practices. If students can be tested, why not their teachers?

Predictable controversy has erupted around this issue. How can you tell a good teacher from a poor one? Teaching is an act that represents planning, curriculum use, knowledge, assessment skills, and classroom management, to name but a few of the functions teachers are responsible for. There are, as well, and no less important to many, the attitudinal goals that teachers address: how to make learning challenging; how to motivate students; how to serve as a model adult who cares for the growth and development of each and every student. In addition, there are the performance requirements that much of earlier research on teacher behavior centered on. The ability to communicate, to be fluent in the standard pedagogical methods, such as leading discussions, conducting a demonstration, delivering a lecture, asking the right questions at the right time, are illustrations of these requirements. Moreover, because teaching is supposed to be instrumental; that is, having been taught, students are supposed to have learned, we must attend to the knowledge base in techniques that

promote achievement. Principles of instruction, including how concepts are best learned, the task of sequencing instructional opportunities, the provision of feedback to students to guide their learning, and the design of practice materials relate to some of the knowledge that is available about instruction.

Observational Approach

A simple minded approach to teacher competency testing would involve the identification of these core tasks and then the careful observation of teachers against particular standards. Among the flaws in such an approach is the lack of agreement on what these essential elements should be as well as a lack of standards to be used in their assessment. Part of the problem here is strictly a weak-knowledge issue. Research on teacher behaviors has been largely correlational in approach. We do not know whether certain behaviors of teachers result in student learning, even though we know that students who learn often have teachers who exhibit particular actions. For example, teachers who ask higher order cognitive questions are thought to produce learning. This relationship may exist, however, because the students in particular classes are able to deal with higher order questions, and create an environment in which this relationship is allowed to flourish. If the knowledge base is insecure in the learning and pedagogical area, it is even less so in the areas of modeling and attitude development, in classroom management skills, in curriculum planning, and in teacher-based assessment. So, for the moment, let us suspend the checklist approach to teacher competency assessment based on classroom observation of certain core skills. If

we were to pursue looking at teachers teach as an approach, we would focus on the interactive portion of what teachers do when they are in classrooms and reduce attention to their preparation or post-instruction analysis. An alternative to teacher observation via structured observation is the holistic judgmental approach, where peers or others in authority observe teachers and decide how well the teacher is doing. Much of the research in performance assessment shows that such approaches are relatively unreliable because they are susceptible to personal views of what good teaching is.....and we have said there is no clear set of precepts. For example, orderly classrooms may or may not be important; there may be a productive kind of disorder, or even a nonproductive kind of order.

Also, at the heart of concerns for observational systems, either structured or very much open-ended, is a concern for individual differences. Teachers exhibit individual differences not only in ability but in their preferences for certain behaviors. Just as a writer may be ultimately effective by carefully outlining before the prose is penned (or processed, these days), another might be just as effective starting out and successively revising as the work is produced.

Outcome Approaches

This analysis leads directly to the issue of the "bottom-line." Perhaps how good a teacher one is shows up best in how well students learn. This approach to teacher assessment has had a long and occasionally productive history. Its motto is to test teachers by testing students; if students learn, then teachers have been

effective. Beyond laboratory experiments in mini-lessons or microteaching, where a particular task is provided for a relatively brief instructional period, 15 minutes to one or two weeks, the problems with assessing teaching by testing students show a number of difficulties. First, and most obvious, is the problem of decising what to test. If the assessment is supposed to be real and a part of the regular curriculum rather than an "experiment," then students may be unready (perhaps because of poor prior instruction) to learn the desired tasks. Some teachers will look bad because of the cumulative history of students in their classrooms. Others may benefit from last year's teacher efforts. So what desired tasks are fair? One approach might be to have tasks identified that are appropriate to learners and judge the teacher's effectiveness on teaching these goals. Such an approach was tried in the Stull Act in California in the early 1970s. A persistent difficulty in this and many other approaches based upon teaching to objectives is the range in difficulty of tasks assessed and the meanings we draw from these. For example, even though solving differential equations is an intellectually more challenging task than solving basic computations, in fact it may be considerably easier when student population characteristics are considered. None of these equating issues would present insurmountable problems were we dealing with groups of teachers and attempting to draw a group estimate of performance. But a central reality is that we are expecting to make decisions about individual teachers, not groups, and the measurement requirements for such decisions are much more stringent.

One apparent way around this issue is to deal with standardized tests, so that everyone is tested on the same content. Differences of

opinion exist about the utility of this approach. Concerns center about the documented lack of correspondence between what is taught in the curriculum and the test content covered. Furthermore, such tests are particularly sensitive to background characteristics of students, such as social class. This is why schools in wealthy areas almost inevitably produce student performance in higher ranges than those schools from poor neighborhoods. Again, even adjusting for these differences, and there is disagreement about how appropriate adjustments are made for individual teachers, teachers still benefit or hindered by the instructional history of students, histories over which the teacher has no control. And how do we treat teachers who teach disabled and gifted students?

Another problem is what to do with unanticipated outcomes, even if we would agree on a reasonable set of outcome measures. The old saw about learning geometry but learning to hate mathematics is the point here. How are the range of acceptable outcomes to be addressed and weighted?

Readiness

Thus far, we have considered looking directly at teaching and looking at a set of desired outcomes of teaching, that is, measured student learning. Additional approaches focus on the teachers' readiness to teach rather than the teaching act or the outcomes of instruction. For instance, what should a teacher know before being permitted to instruct? Certainly, much public agreement is found on the topic of basic skills. We do want our teachers to be correct in the way they use language, perform basic computations, and such - a

minimum competency approach. But this approach is limiting and no one wants minimums to become maximums. What about competence and understanding of subject matter? How much mathematics should teachers know and how well need they know it? What about knowledge of the educational process? The National Teachers Examination, used for certification and published by the Educational Testing Service, has been removed by the publisher from use for master teacher and merit pay decisions. Knowledge about weak knowledge may have even less to offer than demonstration (through observation as discussed earlier) in the matter of competency.

Institutional Effects

One common concern about the assessment of teacher competency relates to the institutional responsibility of the school and the district in which the teacher teaches. To what extent are teachers' needs for curriculum, up-to-date texts, and other support being met? What kind of workplace is the school? Does the principal reward good teaching and high energy? To what extent is competence nested within setting?

Teachers as Employees

Because teachers have organized collectively, certain responses to competency testing may derive from employee/management relationships rather than from the strength of the research base in support of particular options. Among such issues are the identification of "special" teachers for incentive pay structures where the argument is made that the entire salary structure for most

teachers is too low. This issue relates to the ethics of employing teachers during times of need who may not be fully qualified and placing them in jeopardy following a long career commitment. Interrelated to this issue is the idea of tenure, intended to protect teachers from politically inspired decisions.

Marketplace economies have rather different premises, including supply and demand fluctuations, incentive structures, and ways to identify poor performers. With regard to this last point, representatives of teacher organizations have concerns related to the capacity of any performance identification system to provide opportunities to improve rather than simply to weed out low teacher performers, against any criterion.

Social Policy

This litany of issues, questions, and concerns does result in some recommendations. For clearly, social policy is well ahead of the technical base that could be recommended in confidence. One common recommendation is:

LOCATE CONTROLS ON TEACHER ENTRY INTO THE PROFESSION IN CERTIFYING INSTITUTIONS OF HIGHER EDUCATION.

Colleges and universities that offer degrees and provide diplomas must be held accountable for their products. If teachers cannot read adequately or know no mathematics, how can they be graduated and passed along to teacher credentialing programs? And if occasional errors are made, what is the further responsibility of schools of education? Certainly, before certification, teachers must demonstrate capacities in basic skills and subject matter expertise. If not, what

inferences should be drawn about the certifying institution? A second recommendation is:

EMPLOYERS HAVE THE RIGHT TO SET ENTRY STANDARDS; NO ONE CAN FORCE THEIR OWN EMPLOYMENT.

And a corrolary is:

ONCE EMPLOYED, EMPLOYERS HAVE GENERAL OBLIGATIONS AND SPECIFIC CONTRACTUAL RELATIONSHIPS.

Because it is demonstrable that the knowledge base for teaching is yet to be fully developed and that formal measurement feasibly can only assess a portion of important competencies, we are in a situation where policy implementation will proceed but mistakes may be made. Thus, it is critical that the following recommendations be considered:

INVOLVE TEACHERS CLEARLY ALONG WITH OTHER APPROPRIATE PROFESSIONALS IN ANY TEACHER COMPETENCY ASSESSMENT POLICY.

PROVIDE MEANS FOR IMPROVING SPECIFICALLY IDENTIFIED DEFICIENCIES.

EDUCATE ENACTING AUTHORITIES ON THE LIMITS OF EITHER TEACHER TESTING OR STUDENT TESTING (FOR TEACHER ASSESSMENT) WITH REGARD TO BASIC UNRESOLVED PSYCHOMETRIC FACTS THAT IMPINGE ON UTILITY OF THESE APPROACHES.

In the last instance, consider measurement error, for example.

Finally, as a strategy for implementation, in view of the conflicting views and levels of knowledge about the process, consider the final recommendation:

USE A MULTIPLE CRITERION APPROACH IN JUDGING TEACHING.

The chances of serious error are greatly reduced when multiple imperfect options are used.

Research Implications

As always, more research is desirable, and the NIE through its planned Center on Teacher Quality has developed a focus for such research. However, to the extent possible, consideration of technical as well as professional approaches to this complex series of issues is suggsted. Strategies for aggregating information (from multiple criteria), for decomposing performance so as to identify better particular contributions, for improving measurement and placing it (at the colleges and universities) where it can do most good, must be pursued. The issue will not disappear, so our methods for dealing with it must improve.

Note

These comments were derived from the literature, conversations, and conferences with representatives from the following organizations:

The National Education Association - Instruction and Professional Development

The American Federation of Teachers

The Interinstitutional Conference of the American Educational Research Association

The Pennsylvania School Study Council

The U.S. Department of Defense, Dependent Schools-Evaluation Division

The Council of Chief State School Officers

The Office of Technology Assessment

No endorsement of these remards in whole or in part by any of these organizations is to be inferred.

Future R&D Needs in Educational Testing and Evaluation

The quality of this nation's educational system has been the subject of considerable scrutiny and criticism in the past several years. Education in the United States has been examined and judged to be seriously wanting, so much so that the nation is said to be "at risk," a conclusion drawn not only by the National Commission on Excellence in Education, but in varying degrees by the National Science Board, the Carnegie Report, the Education Commission of the States, and the Twenty Century Fund, to name but a few of the recent reports. In fact, the large number of reports on the state of education in the United States has led the Education Commission of the States (ECS, 1983) to characterize 1983 as the "Year of the Report on Education."

Educational testing and evaluation has played an important role in making judments about the educational enterprise and likewise has been expected to play a significant part in its improvement. Determinations about the mediocrity -or worse- of the current system, for example, are based frequently on test results. The decline in scores on the Scholastic Aptitude Test and the discouraging implications of international comparisons of student achievement have contributed substantially to current quality judgments. The tremendous media attention and public interest accorded the school-by-school results of state assessments and other standardized tests further substantiate the role of tests in documenting educational accomplishments.

The role of testing and evaluation in school improvement efforts has also been visible. Minimum competency testing programs, for instance, have been offered as one solution for improving and assuring quality education; teacher competency testing represents another instance of a similar phenomena. By setting standards and administering tests, the quality of the enterprise is supposed to be strengthened. The effective schools literature too points to the role of tests in improved school performance, through continual monitoring and assessment of student progress, through the expectations they imply, and through the standards they exemplify.

More generally, evaluation has also been thought to have a strong role in promoting school quality, not only by facilitating accountability but by fostering analysis of the strengths and weaknesses within schools and districts and by stimulating corrective actions. Formative evaluation systems which can help schools analyze their context, process, and a range of outcomes have been the subject of recent study.

Given the testing and evaluation's role in judging and promoting educational quality, what are the implications for research and development? A number of researchers gathered together to explore this question, including Gene Glass, Lorrie Sheppard, Robert Linn, Ernest House, Robert Stake, Mary Lee Smith, Eva Baker, Leigh Burstein, and Joan Herman. The issues they raised are summarized next.

Validity

The use of testing and evaluation in making judgments about educational quality and in promoting school improvement assumes the validity of the process and its results. Few concepts, in fact, are as fundamental to educational testing and evaluation as is the concept of validity. It guides the construction of tests and the conduct of evaluation studies to a considerable degree. There is evidence to suggest, however, that the concept, as it has evolved, is in need of serious revision.

Most generally, validity refers to the correctness or appropriateness of an inference about something. "Validity attaches to a conclusion..." Cronbach says, pointing out that many conclusions can be drawn from the same data and not all conclusions will be equally warranted (Cronbach, 1982, p. 106). Similarly, Cook and Campbell (1979) say that validity refers to the "...truth or falsity of propositions, including propositions about cause." Although we speak about valid experiments or valid tests, what we mean is that the experiments, evaluations, or tests are constructed in such a way that we can draw valid conclusions from them.

The traditional explication of the validity concept is by Campbell and Stanley (1963) and Cook and Campbell (1979). More recently, Cronbach (1982) has challenged this traditional view with a reformulation of the concept of validity itself and of the issues involved, especially as applied of evaluation. Cronbach (1980) has also extended and revised his notion of test validity, which one might view as a special case of the overall concept. Among the important observations he suggests is that only a particular interpretation of a

test can be validated, not the test itself.

Cronbach's formulation changes the definitions of both internal and external validity and shifts the importance accorded to each. What are its implications for how tests should be constructed and evaluation studies conducted? A potential line of research would pursue the reformulation and see if these notions could be tested, elaborated, and revised. Furthermore, one might explore potential changes in the technology of evaluation and testing which would be entailed if such a different notion of validity were widely accepted. Research is needed, in addition, to examine its implications for other evaluation approaches, such as qualitative approaches.

Needs for Different Kinds of Testing

Despite the already pervasive use of tests, there is a need for new kinds of testing. The new kind of testing that is needed differs from the majority of existing testing in a number of different needs. First, the focus on excellence requires tests of higher order skills, the ability to use content knowledge to solve problems rather than tests which stress minimums. Second, local and school variability in content and curricular goals require tests that match both the common and unique goals of schools rather than just the least common denominator of content that is often stressed in current standardized tests. Third, using tests to help guide instruction and adapt it to student and group needs requires tests that provide diagnostic information to students and teachers rather than just a global score showing a student's standing relative to a national norm.

The design of useful diagnostic tests may represent the most

difficult problem, reflecting the interaction of important content and methodological issues. First, there has been a lack of strong theory to guide the selection of test content. Although recent developments in cognitive psychology and artificial intelligence may provide some guidance, the development of conceptualizations which are generalizable and sensitive to the nature of students' misconceptions requires considerable work.

Second, existing psychometric theory, which has considerable power for some purposes, is not well suited to diagnostic tests. Neither classical nor item response theory is designed to deal with diagnostic testing problems. Both approaches rely on an assumption of unidimensionality and treat deviations as noise. Yet, it is precisely those deviations from a single dominant dimension that are of central concern in diagnostic testing. Substantial research to develop or adapt new theories and strategies is needed.

In addition, diagnosis requires a level of detailed information that may be too time consuming to collect and too difficult for a teacher to maintain and use with existing paper-and-pencil testing technology. While the wide availability of microcomputers in schools will assuage this problem, generalizable and flexible approaches are needed that can be adapted to a range of curriculum and content areas with their unique structures and for an array of item types. Likewise, there needs to be research to assure the usefulness of the solution for teachers in their classrooms.

While diagnostic testing information provides information at the individual level for assessing individual student needs, it also can provide information for assessing strengths and weaknesses and

instructional progress within classroom groups or for the class as a whole. Likewise, it should be possible to use tests which assess students' classroom progress to make judgments about instructional effectiveness at the school and district levels. Such multipurpose testing deviates from current practices which feature overlapping testing requirements for each decisionmaking need. While desires for cost efficiency, for conserving instructional time and for increasing the sensitivity of testing programs to classroom needs and problems argue for such an approach, a number of methodological problems remain to be solved.

Comprehensive Information Systems

Test results represent only one kind of information which is needed to evaluate and understand what is going on in schools and to stimulate their improvement. If it is to be useful, schools need information which is sensitive to their local needs and which represents issues and concerns of particular local interest. They likewise need information about school context and instructional and school processes as well as a range of school outcomes if they are to analyze and make sense out of their environment, determine strengths and weaknesses, explore potential cause and effect relationships, and make plans for improvement.

Districts, too, need a more comprehensive information base to make judgments about their schools' effectiveness, to plan wisely and to allocate resources most effectively. Information about school and instructional practices, for example, will provide them with another view of how schools are operating and may suggest areas where staff

development is needed. Information about school context, as another example, can help increase the validity of judgments about school effectiveness (e.g., information about students' SES may moderate judgments about the meaning of particular levels of test performance). The examination of context, process and performance over time, furthermore, may uncover trends which are masked in single year analyses but which have important implications for evaluative judgments and/or which signal the need for action.

While comprehensive evaluation systems which serve the needs of local teachers, schools, districts and beyond are theoretically possible, a number of research issues remain to be solved. Among these are how to aggregate information for decisionmaking at various levels, e.g., combining data from various sources to arrive at quality indices; how to devise systems which provide information which is appropriate for policy-making but which is suitably sensitive local goals and emphases and how to promote use at the various levels.

The Effects of Testing

While we work to increase the validity and usefulness of testing and evaluation, there is a need to study the effects of current practices. Although standardized achievement testing threatens to expand to incredible proportions, there are only a few in-depth studies of the effects of such testing on schools, curricula, teachers and students. There has been little in-depth attention to how tests are used to make educational decisions and how test-based information fits in the larger context of policy formation. For the most part it has been assumed that tests are beneficial and need only to be made

better, with little concern for potential side effects.

Ordinarily, testing is viewed as a means of measuring the outcomes of instruction. But there is much folklore and some evidence that testing has its own effects. These effects need to be studied because tests may be an instructional treatment in their own right or because the presence of these effects undermines the validity of test results as an index of learning outcomes. The reactivity of testing, particularly criterion-referenced and minimum competency testing as it is currently practiced, needs scrutiny, particularly in terms of claimed effects on narrowing and simplifying the curriculum.

At the system level, one presumed effect of testing is that test content shapes curriculum. But, we do not know much about whether this happens, how it happens, or the features of the testing program that produce the greatest effects. The examination of the effects of minimum competency testing programs on teachers, students, and curricula represents an interesting case in point, one with significant current policy implications. What are the effects of such programs on the curriculum? What are the effects of such programs on instruction, or remediation, on student competence and what is the meaning of test gains in such programs. Are there different effects in programs where tests are used for grade to grade promotion versus those which are used for high school graduation only?

At the individual level, we know that frequent testing is related to instructional gains. However, their effects are only vaguely understood. Tests as the "cause" of increased learning could occur because they increase motivation (to study, to pay attention, etc.), because they target attention, because feedback on correct and

incorrect responses aids learning, or because the student becomes more efficient in learning only for the tests, to name a few possibilities. This last explanation raises important questions about the validity of the test as a proxy for a larger intended content domain and about its actual effects in long term significant learning.

Cost-efficient Analyses for Educational Policy-making

The public and its policymakers have focused increasing attention on statistical indices of school performance, seeking periodic large scale assessments of student progress and other quality indicators. These assessments have adhered to traditional standards of statistical methodology and educational measurement, involving preordinate specification of subjects to be assessed, focusing on measurement of a few popular topics, and elaborate sample frameworks for achieving representative pictures of a population, followed by extensive coordination, data collection and analysis efforts. The methodology, while refined, has been expensive, sometimes slow in producing its findings, and often narrow in scope.

Meanwhile, there exists no shortage of data about the performance of public schools. Data lie about in file cabinets and computer files. These data are potentially useful as indicators if they can be focused to transform them from parochial and episodic snapshots of educational performance into more representative and consistent information. Can the social indicator information on education derived from expensive, large-scale assessments be obtained at greatly reduced cost from smaller, less representative data files which already exist and which were designed to serve other purposes? What

kinds of social indicator information can be derived from such sources as the files of standardized test companies, State Education Agencies, Local Education Agencies and research projects such as the National Longitudinal Study, the Sustaining Effects Study, and the like? Can such information on trends in educational performance substitute, after appropriate reworking -- for data purchased at higher cost under much less flexible circumstances by preplanned, nationally representative longitudinal assessments?

If appropriate methodologies can be devised, the results will serve not only cost efficiency and goals in large scale assessment, but, also may ultimately serve validity and local decision-making as well. If we can find ways to structure aggregate data from different sources, it may be possible to combine data that reflect local curriculum and its unique emphases for broader decisionmaking and assessment purposes. Building from the school and classroom level out, it might be possible to design multipurpose, locally sensitive information systems that are useful for a variety of decisions at the class, school, district, and higher policymaking levels.